

# Single-channel speech enhancement by PSO-GSA with harmonic regeneration noise reduction

Kalpana Ghorpade<sup>1</sup>, Arti Khaparde<sup>2</sup>

<sup>1</sup>MKSSS's Cummins College of Engineering for Women, Pune, India

<sup>2</sup>Department of Electronics and Communication Engineering, Vishwanath Karad MIT World Peace University, Pune, India

## Article Info

### Article history:

Received Nov 30, 2022

Revised Dec 22, 2022

Accepted May 2, 2023

### Keywords:

Evolutionary algorithm

Harmonic regeneration

Noise reduction

Optimization

Speech intelligibility

## ABSTRACT

Speech quality significantly affects the performance of speech dependent systems. Noise in the background lowers the clarity and intelligibility of speech. The augmentation of speech can increase its quality. We propose a single-channel speech improvement framework that combines particle swarm optimization (PSO), gravitational search algorithm (GSA), and harmonic regeneration noise reduction (HRNR) to minimize speech signal noise and increase speech intelligibility. The proposed hybrid algorithm optimizes the amount of overlap between the noisy speech frames. This helps in reducing the overlapped noise. Then HRNR algorithm is applied to retain the speech harmonics. The algorithm gives improvement in the speech intelligibility for babble, car and exhibition noise. The segmental signal to noise ratio (SNR) is also improved for these noise types. There is improvement in speech intelligibility with minimal speech distortion.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Kalpana Ghorpade

MKSSS's Cummins College of Engineering for Women

Karve Nagar, Pune, Maharashtra, India

Email: kalpana.joshi@cumminscollege.in

## 1. INTRODUCTION

Speech signal gets contaminated by background noise affecting its clarity and intelligibility. In the speech dependent programs namely automatic speech recognition, speaker recognition, speech operated applications quality of speech is very important. These systems give better accuracy if the input speech is with minimum noise. In real environment, background noise, either stationary or non-stationary is everywhere. It gets added with speech signal giving noisy speech. As the additive noise appears in different forms and shapes, it is impossible to track and remove it completely [1]. But it is possible to reduce it. Speech enhancement systems can reduce additive noise present in noisy speech. This improves speech quality which in turn improves the accuracy and efficiency of speech-dependent systems. Besides, improvement in speech intelligibility is crucial in automatic speech recognising systems. Conventional speech advancement method never enhance speech intelligibility in a highly noisy environment [2]. The main challenge in speech improvement is to minimise noise in speech without distorting the speech content and still giving improvement in speech intelligibility.

Dual channel speech enhancement systems give better performance but with extra cost. Neural network-based speech enhancement methods require clean speech to train the system. Single-channel speech enhancement system is quite tough to implement as there is no noise reference. This research aspires to implement a single channel speech improvement system with a hybrid evolutionary algorithm that can boost voice intelligibility without speech distortion. This implementation does not have the reference of background noise or clean speech.

The contribution of this research is: i) implementing combination of PSO and GSA to have a hybrid algorithm so that the advantages of both the algorithms can be used; ii) deciding the fitness function for the modified algorithm; iii) using this hybrid algorithm, decide the percentage of overlap between the noisy speech frames before applying TSNR and harmonic regeneration noise reduction (HRNR); iv) testing the proposed methodology for various kinds of additive noises with various SNR levels; and v) estimating the performance measures for each noise type and contrast the outcomes with common speech enhancement methods. This paper is arranged as follows: papers related to previous research are described in section 2. section 3 gives materials and method; section 4 depicts the implementation outcomes and section 5 gives conclusion.

## 2. LITERATURE REVIEW

The existing speech enhancement methods are classified as spectral subtractive method, statistical-model-based method and subspace method. Spectral subtraction for speech denoising suggested by [2] suffers from musical noise. To reduce musical noise, modulation frequency domain is recommended in literature. Spectral subtraction is carried out in modulation frequency domain [3]. Model oriented speech advancement is executed [4]. Implementing a Kalman filter in modulation domain to predict the spectral amplitudes of speech and noise, model-based speech enhancement is accomplished in modulation domain [5]. According to Buragohain *et al.* [6], sinusoidal modelling is used for speech analysis. A combination of convolutional and recurrent neural network is used for single channel speech enhancement [7]. Wiener filter is the optimal complex discrete fourier transform (DFT) coefficients estimator [8] which does not estimate the optimal spectral magnitudes. Several optimal methods are proposed in literature to obtain spectral amplitudes from noisy observations as it has the importance respecting speech intelligibility and quality [8]. Minimum mean square error (MMSE) spectral amplitude estimator is one of such kind. It requires a priori signal to noise ratio (SNR) estimation. The estimation of a priori SNR is frequently performed using the decision-directed (DD) technique. At low SNR, the DD method overestimates the true value of a priori SNR, which suppresses the speech signal [9]. An unique statistical model that takes into account the time-correlation between succeeding speech spectrum components is described [9]. The two-step noise reduction (TSNR) technique, which addresses the aforementioned issues while retaining the benefits of the DD approach, is suggested [10]. Here, a second step is employed to eliminate the DD approach's bias, which allows for more precise calculation of the a priori SNR. All short-time noise reduction techniques, including TSNR for tiny SNR, produce harmonic distortion in enhanced speech [10].

The short-time noise reduction methods fail to boost speech harmonics at low SNR levels was presented by Plapous *et al.* [11]. They have implemented a new method in which the harmonic characteristic of speech is considered. Processing of the output of a noise reduction method is again done to regenerate the missing harmonics of speech. This method is combined with deep neural network (DNN) for single channel speech improvement [12]. For noise reduction in speech, evolutionary algorithms can be used as these algorithms are not dependent on system structure, they utilize only fitness function. They provide excellent solutions for rough, discontinuous, and multimodal surfaces [13]. Particle swarm optimization (PSO) chooses the search area for the subsequent iteration based on the knowledge currently available about the search space. PSO's exploitative character can lead to premature convergence at local optima, but it can also help solve the problem [14]. So, different modifications are suggested in standard PSO giving modified PSO (MPSO) [15], accelerated PSO (APSO) [16], craziness based PSO (CRPSO) [17] to improve the efficiency. According to Kunche *et al.* [18], a blend of PSO and gravitational search algorithm (hybrid PSOGSA) is proposed to improve output SNR compared to input. A combination of spectral filtering MMSE and PSO, (MMSEPSO) is suggested [19].

## 3. MATERIALS AND METHOD

### 3.1. Particle swarm optimization

The PSO is introduced [20]. Here the particles are nothing but the possible solutions to the problem. They are initialized randomly at in the start [20]. Position of a particle (i) in the t th iteration is represented by vector  $u_i$  and its velocity is given by  $v_i$ .

$$v_i^t = W * v_i^{t-1} + C_1 * rand_1(p_i - u_i^{t-1}) + C_2 * rand_2(p_g - u_i^{t-1}) \quad (1)$$

$$u_i^t = u_i^{t-1} + v_i^t \quad (2)$$

where  $C_1$  and  $C_2$  are random numbers, W is the inertia. It is the inclination of the particle to move in the direction as it moved in the last iteration. In each iteration, the objective function value is estimated for every

particle. Based on this value, the best particle of that iteration (global best particle) is decided which has the best value. If the problem is of maximization, then the best particle will be with maximum value of objective function. Global best position ( $p_g$ ) refers to the location of the best particle of the swarm. To determine the local best value of a particle that gives the local best location ( $p_i$ ), the particle's current fitness value is compared to its fitness value from the previous iteration. In (1) provides the particle's velocity.

Depending on the velocity, the particle positions are modified as per (2) for the next iteration. The steps get repeated, and the algorithm runs through the iterations until the stopping criterion is satisfied. The algorithm gives an optimal solution to the problem.

### 3.2. Gravitational search algorithm

Rashedi *et al.* [21] suggested a new high performance heuristic algorithm named as GSA formulated on the gravitational law. Here, objects are agents. The measurement of the performance of the agents is based on the masses of the agents. The heavier mass means the better solution. The force of gravity causes objects to be drawn together. All the objects attract each other and get moved towards the objects with heavier masses due to this force. The heavy masses move slowly. The location of the mass exemplifies the achievement of a solution to the issue [21]. For N agent system,  $X_m = (x_m^1, x_m^2, \dots, x_m^d, \dots, x_m^n)$  depicts the position of the mth agent where  $m=1; 2; \dots; N$ .

The force on mth mass due to kth mass at t is given by (3):

$$F_{mk}^d(t) = G(t) \frac{M_m(t) * M_k(t)}{R_{mk}(t) + \varepsilon} (x_k^d(t) - x_m^d(t)) \quad (3)$$

where  $M_k$  and  $M_m$  are the gravitational masses of agent k and agent m respectively,  $x_m^d$  and  $x_k^d$  represent the position of mth agent and position of k th agent respectively in dth dimension,  $G(t)$  is a gravitational constant at time t,  $R_{mk}(t)$  is the Euclidian distance among the two agents m, k.

$$R_{mk}(t) = \|X_m(t), X_k(t)\|_2 \quad (4)$$

In (5) gives the total force acting on m in a dth dimension.

$$F_{md}(t) = \sum_{k=1, k \neq m}^N rand_k F_{mk}^d(t) \quad (5)$$

where  $rand_k$  is a random number in the interval [0,1]. In (6) gives the acceleration of the agent m, at t.

$$a_m^d(t) = \frac{F_{md}^d(t)}{M_{mm}(t)} \quad (6)$$

where  $M_{mm}(t)$  is the inertial mass of mth agent. The velocity and position of m are given by (7) and (8) respectively.

$$v_m^d(t+1) = rand_m v_m^d(t) + a_m^d(t) \quad (7)$$

$$x_m^d(t+1) = x_m^d(t) + v_m^d(t+1) \quad (8)$$

where  $rand_m$  is a standardised random variable in the interval [0,1]. G is initialized at the start and goes on reducing with time for controlling the searching accuracy.

$$G = G_0 e^{-s \frac{Iter}{Iterations}} \quad (9)$$

here  $G_0$  is the initial value of G and s are the damping constant, iter is the current iteration number and the total count of iterations set by the technique is known as iterations. The masses are calculated by using (10) and (11):

$$m_m(t) = \frac{fit_m(t) - worst(t)}{best(t) - worst(t)} \quad (10)$$

$$M_m(t) = \frac{m_m(t)}{\sum_{k=1}^N m_k(t)} \quad (11)$$

where  $fit_m(t)$  represents the fitness value of the agent m at time t,  $worst(t)$  and  $best(t)$  are the minimum fitness value at t and the maximum fitness value at t respectively when the problem is of maximization,  $m_i(t)$  is an intermediate variable in particle mass calculation.

### 3.3. Proposed algorithm

In conventional PSO, potential solution to the problem is found with the use of exploitative nature but sometimes the solution may get trapped into the local optima [20]. In GSA, the particles have no memory ability since just the current position information is used in the updating process during the assessment [22]. By combining it with PSO we include memory to the particles. So, we implemented the combination of GSA and PSO, the hybrid algorithm which can overcome the shortcomings of the two algorithms [22]. We used this algorithm as a pre-processing step for the combination of TSNR and HRNR standard method. It proved to be beneficial in improving the intelligibility and the clarity of the enhanced speech.

Initially the noisy speech is framed with 25 ms duration and windowed using hanning window. Then 400-point FFT is calculated for each frame. To decide the overlap percentage within the speech frames, we used PSOGSA. The parameters of PSOGSA such as the number of agents (N), total number of iterations (iterations), dimension of the problem space (d), constants C1, C2, and G0 are initialised. Instead of randomly generating the agents, we have considered them as the samples of the speech signal. So, the amount of agents is same as the number of samples in speech signal. The velocity of the agents is initialised randomly. Current position of the agents is the function of the sample values of the speech signal. The upper bound and the lower bound for the agent position are decided at the start based on the maximum and minimum value in the speech signal. In every iteration, based on the current position, the fitness function (objective function) of each individual is estimated. The algorithm finds the maximal value of the fitness function which is the global best of that iteration. The corresponding position of the global best gives the global best position. The maximum value of fitness function is considered as the finest fitness and the minimum value is considered as the worst fitness. G is updated as per (9). For every agent, mass is estimated based on the current fitness of the agent, best and worst value of the fitness as per (10) and (11) and the force and acceleration are estimated as per (5) and (6) respectively. The new velocity is found for each agent according to (12). The velocity equation is modified to have the combination of PSO and GSA. According to the new velocity, position of the particles is updated by (13) while entering the new iteration. The algorithm goes through all the iterations to find the optimized global best value.

$$v_m^d(t+1) = rand * velocity(t) + C_1 * rand * (acceleration(t)) + C_2 * rand * (gbest(t) - x_m^d(t)) \quad (12)$$

$$x_m^d(t+1) = x_m^d(t) + v_m^d(t+1) \quad (13)$$

Figure 1 gives the block diagram of the algorithm. Using DD approach a priori SNR is estimated. Then standard TSNR algorithm with harmonic regeneration and noise reduction is applied to get the enhanced speech [8], [9].

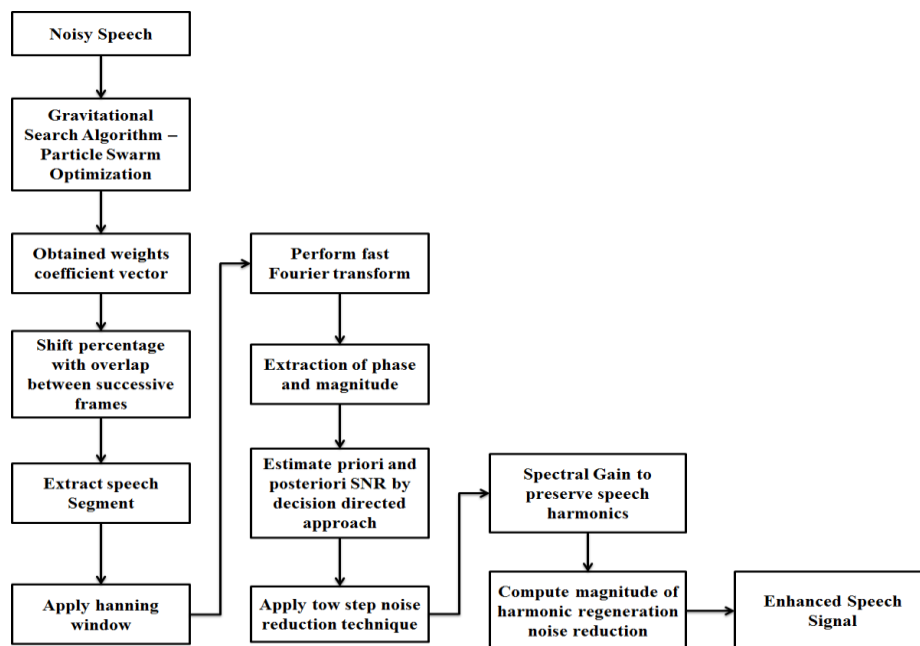


Figure 1. Block diagram of the proposed framework

### 3.3.1. Parameter selection

As the number of agents (N) is large in our case, we selected lesser number of iterations which is equal to 20. G0 is kept equal to 1 and s is equal to 23. The acceleration coefficients C1 and C2 are 0.5 and 1.5 respectively as  $C1 + C2 \leq 4$  [20]. Table 1 gives the parameters of the proposed framework.

Table 1. Parameters of proposed algorithm

Parameter	Value
N	No. of speech samples in noisy speech
d	1
Iterations	20
G0	1
C1	0.5
C2	1.5
s	23

### 3.3.2. Objective function

Here, the particles or agents are nothing but the samples of the input noisy speech sentence. Using it, the linear prediction coefficients are found for the filter and the signal is passed through the designed filter. The filter output is considered as the fitness function value for individual particle. In (14) gives the fitness function.

$$a = lpc(x_i(t), 3) \quad (14)$$

$$fitness = filter[0 - a(2:end)], 1, x_i(t)]$$

The algorithm finds the maximum value of the fitness function in each iteration. The most optimized value of the fitness function is found by the algorithm in twenty iterations. Using this value, the frame overlapping percentage is calculated and is used while applying TSNR algorithm. Instead of keeping the frame overlap percentage fixed, we used the parameter optimized by the PSO-GSA which in turn gives the amount of overlap. It reduces reverberation effect in the enhanced speech.

### 3.3.3. Stopping criterion

We kept the total number of iterations equal to 20. We get optimization in first 6 to 7 iterations. The algorithm stops executing when it passes through all the iterations giving optimized value for the global best. The most optimized value of the fitness function given by the framework is used to decide the overlap amount within the noisy frames. Speech harmonics are maintained at low SNR levels while noise is removed using a TSNR algorithm with HRNR.

## 4. RESULTS

The proposed algorithm's performance is evaluated using the developed noisy speech corpus (NOIZEUS) database [23]. The objective speech quality metrics are perceptual evaluation of speech quality (PESQ) [24] and segmental SNR [25]. Sentences with additive noises of three different sorts and three SNR levels are obtained from the NOIZEUS database. We used babble, car, and exhibition noises. The results are compared the results of MMSE-PSO [19], log-MMSE, TSNR, and TSNR with HRNR algorithms [11]. The PESQ value for various approaches is shown in Table 2. It demonstrates that the suggested algorithm improves PESQ more than the TSNR and HRNR algorithms. It has been found that the PESQ improvement for 0dB is greater than for 5 and 10 dB. The improvement in segmental SNR is analyzed in Table 3. The proposed framework gives the same amount of improvement in segmental SNR with increase in PESQ as compared to TSNR and TSNR with HRNR.

Figure 2 displays a PESQ comparison for the exhibition noise which shows that the suggested algorithm performs better at 0 dB and at 5 dB than 10 dB and compared with other algorithms. Figure 3 gives the spectrogram of clean speech, noisy speech and enhanced speech by a female speaker with 5 dB car noise. The spectrogram shows that the noise is reduced in the enhanced speech telling the quality improvement. With the noise reduced, the output has clear formants. In six iterations, the method converges and produces optimum fitness function.

Table 2. Output PESQ				
Noise type	Method	0 dB	5 dB	10 dB
Babble	MMSE-PSO	1.75	1.895	2.45
	Log MMSE	1.8630	2.2065	2.5973
	TSNR	2.55	2.77	3.0971
	TSNR+HRNR	2.4513	2.70	3.081
	Proposed algorithm	2.6930	2.7756	3.1034
Car	MMSE-PSO	1.75	2.13	2.25
	Log MMSE	1.9221	2.3394	2.7090
	TSNR	2.3271	2.6195	2.9049
	TSNR+HRNR	2.3286	2.6430	2.9217
	Proposed algorithm	2.4209	2.6712	2.9806
Exhibition	MMSE-PSO	1.75	1.95	2.4
	Log MMSE	1.6712	2.1703	2.5334
	TSNR	2.3287	2.5568	2.9188
	TSNR+HRNR	2.3586	2.5738	2.9349
	Proposed algorithm	2.4918	2.5868	2.9604

Table 3. Improvement in seg SNR				
Noise type (dB)	TSNR	TSNR+HRNR	Proposed algorithm	
Babble	0	7.46	7.47	7.5
	5	4.12	4.15	4.73
	10	3.62	3.73	3.77
Car	0	8.06	7.99	8.16
	5	6.32	6.23	6.59
	10	3.95	3.86	4.17
Exhibition	0	7.46	7.46	7.49
	5	5.70	5.78	5.78
	10	3.73	3.75	3.76

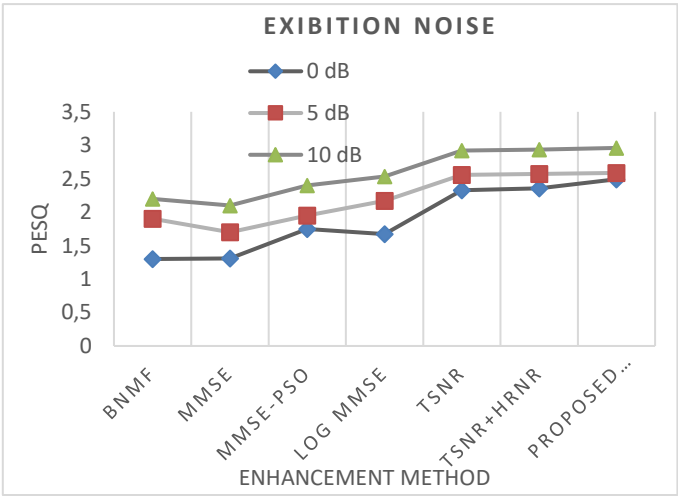


Figure 2. PESQ for exhibition noise

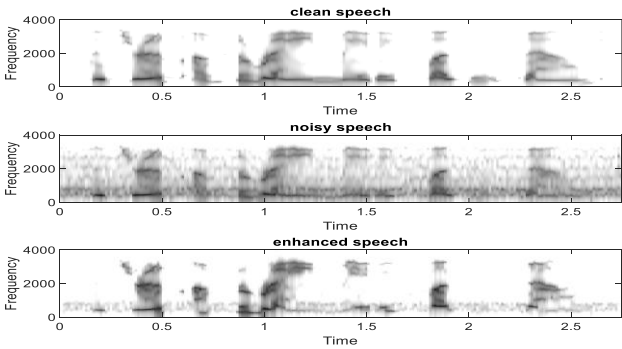


Figure 3. Spectrogram of Sp12 ‘The drip of the rain made a pleasant sound’ (female speaker) with 5 dB car noise PESQ of enhanced speech=2.6790




## 5. CONCLUSION

We used a single-channel speech enhancement system in our work. In the suggested technique, we utilised PSO-GSA to determine how much overlap there was between the noisy voice frames before using a TSNR algorithm with harmonic regeneration. This method results in an increase in PESQ, indicating an improvement in the comprehension of input speech. For all three noise classes, this improvement is more pronounced for 0 dB noise. Compared to car noise, we saw a greater level of improvement in PESQ for babbling and exhibition noise. There is increase in segmental SNR value indicating the quality improvement of the enhanced speech. Reduction in noise is also clear from the spectrograms. Algorithm converges within 10 iterations. Hybrid algorithms such as the combination of evolutionary algorithms may be employed to optimize the parameters of existing speech enhancement algorithms for improving the intelligibility.




## REFERENCES

- [1] A. M. Kondo, *Digital speech: coding for low bit rate communication systems*, 2nd ed. Chichester: John Wiley & Sons, 2005.
- [2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, Apr. 1979, doi: 10.1109/TASSP.1979.1163209.
- [3] K. Paliwal, K. Wójcicki, and B. Scherer, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Communication*, vol. 52, no. 5, pp. 450–475, May 2010, doi: 10.1016/j.specom.2010.02.004.
- [4] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proceedings of the IEEE*, vol. 80, no. 10, pp. 1526–1555, 1992, doi: 10.1109/5.168664.
- [5] Y. Wang and M. Brookes, "Model-based speech enhancement in the modulation domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 580–594, Mar. 2018, doi: 10.1109/TASLP.2017.2786863.
- [6] R. Buragohain, R. A. Reddy, Y. Venkatesh, G. A. Prabhakar, and C. V. R. Rao, "Single channel speech enhancement using masking based on sinusoidal modeling," in *Recent Trends in Image Processing and Pattern Recognition: 4th International Conference, RTIP2R 2021*, 2022, pp. 330–337, doi: 10.1007/978-3-031-07005-1\_28.
- [7] J. Li, H. Zhang, X. Zhang, and C. Li, "Single channel speech enhancement using temporal convolutional recurrent neural networks," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Nov. 2019, pp. 896–900, doi: 10.1109/APSIPAASC47483.2019.9023013.
- [8] P. C. Loizou, *Speech enhancement: theory and practice*. Boca Raton, FL: CRC Press, 2013.
- [9] I. Cohen, "On the decision-directed estimation approach of Ephraim and Malah," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, pp. 293–296, doi: 10.1109/ICASSP.2004.1325980.
- [10] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 6, pp. 2098–2108, Nov. 2006, doi: 10.1109/TASL.2006.872621.
- [11] C. Plapous, C. Marro, and P. Scalart, "Speech enhancement using harmonic regeneration," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, pp. 157–160, doi: 10.1109/ICASSP.2005.1415074.
- [12] N. Jamal, N. Fuad, and M. Shaabani, "A hybrid approach for single channel speech enhancement using deep neural network and harmonic regeneration noise reduction," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 10, pp. 243–248, 2020, doi: 10.14569/IJACSA.2020.0111033.
- [13] D. J. Krusienski and W. K. Jenkins, "Adaptive filtering via particle swarm optimization," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, 2003, pp. 571–575, doi: 10.1109/ACSSC.2003.1291975.
- [14] A. Aggarwal, T. K. Rawat, and D. K. Upadhyay, "Design of optimal digital FIR filters using evolutionary and swarm optimization techniques," *AEU - International Journal of Electronics and Communications*, vol. 70, no. 4, pp. 373–385, Apr. 2016, doi: 10.1016/j.aue.2015.12.012.
- [15] D. J. Krusienski and W. K. Jenkins, "Particle swarm optimization for adaptive IIR filter structures," in *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753)*, 2004, pp. 965–970, doi: 10.1109/CEC.2004.1330966.
- [16] X.-S. Yang, *Nature-inspired metaheuristic algorithms*. Frome United Kingdom: Luniver Press, 2008.
- [17] S. Mandal, S. P. Ghoshal, R. Kar, and D. Mandal, "Design of optimal linear phase FIR high pass filter using craziness based particle swarm optimization technique," *Journal of King Saud University - Computer and Information Sciences*, vol. 24, no. 1, pp. 83–92, Jan. 2012, doi: 10.1016/j.jksuci.2011.10.007.
- [18] P. Kunche, G. S. B. Rao, K. V. V. S. Reddy, and R. U. Maheswari, "A new approach to dual channel speech enhancement based on hybrid PSO-GSA," *International Journal of Speech Technology*, vol. 18, no. 1, pp. 45–56, Mar. 2015, doi: 10.1007/s10772-014-9245-5.
- [19] R. S. Selvi and G. R. Suresh, "Hybridization of spectral filtering with particle swarm optimization for speech signal enhancement," *International Journal of Speech Technology*, vol. 19, no. 1, pp. 19–31, Mar. 2016, doi: 10.1007/s10772-015-9317-1.
- [20] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.
- [21] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: a gravitational search algorithm," *Information Sciences*, vol. 179, no. 13, pp. 2232–2248, Jun. 2009, doi: 10.1016/j.ins.2009.03.004.
- [22] S. Mirjalili and S. Z. M. Hashim, "A new hybrid PSO-GSA algorithm for function optimization," *International Conference on Computer and Information Application*, Tianjin, China, 2010, pp. 374–377, doi: 10.1109/ICCIA.2010.6141614.
- [23] Y. Hu and P. Loizou, "Subjective evaluation and comparison of speech enhancement algorithms," *Speech communication*, vol. 49, pp. 588–601, 2007, [Online]. Available: <https://ci.nii.ac.jp/naid/80018409126/>
- [24] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, 2001, pp. 749–752, doi: 10.1109/ICASSP.2001.941023.
- [25] P. Mermelstein, "Evaluation of a segmental SNR measure as an indicator of the quality of ADPCM coded speech," *The Journal of the Acoustical Society of America*, vol. 66, no. 6, pp. 1664–1667, Dec. 1979, doi: 10.1121/1.383638.

**BIOGRAPHIES OF AUTHORS**

**Kalpana Ghorpade**    has received masters degree in Electronics Engineering in 2010. She has an experience of 21 years in academics. Currently she is working as an assistant professor at Cummins college of engineering for women, Pune, Maharashtra, India. She is pursuing Ph.D. in speech enhancement domain. Her areas of interest are signal processing, speech processing, and speech enhancement. She can be contacted at email: kalpana.joshi@cumminscollege.in.



**Arti Khaparde**    is a Professor at the Department of Electronics and Communication Engineering, MIT World Peace University, Pune, Maharashtra, India. She has received Doctor of Philosophy in the Faculty of Electronics and Communication Engineering from JNTU, Hyderabad, India in 2009. She has total 19 years of experience in academics and research. She has publications in several reputed national, and international journals. There are 3 patents in her name. Her areas of interest are speech processing, video processing, computer vision, and artificial neural networks. She can be contacted at email: arti.khaparde@mitwpu.edu.in.